Technical Report 1084

An Approach to Evaluating Distance Learning Events

Robert A. Wisher
U.S. Army Research Institute

Christina K. Curnow
George Washington University
Consortium Research Fellows Program

June 1998

United States Army Research Institute for the Behavioral and Social Sciences

U.S. Army Research Institute
for the Behavioral and Social Sciences

A Directorate of the U.S. Total Army Personnel Command

EDGAR M. JOHNSON
Director

Technical Review by

Ruth Phelps
Angelo Mirabella

NOTICES

DISTRIBUTION: Primary distribution of this Technical Report has been made by ARI. Please address correspondence concerning distribution of reports to: U.S. Army Research Institute for the Behavioral and Social Sciences, Attn: TAPC-ARI-PO, 5001 Eisenhower Ave., Alexandria, VA 22333-5600.

FINAL DISPOSITION: This Technical Report may be destroyed when it is no longer needed. Please do not return it to the U.S. Army Research for the Behavioral and Social Sciences.

NOTE: The findings in this Technical Report are not to be construed as an official Department of the Army position, unless so designated by other authorized documents.

REPORT DOCUMENTATION PAGE

Form Approved OMB No. 0704-0188

1. AGENCY USE ONLY *(Leave Blank)*	2. REPORT DATE	3. REPORT TYPE AND DATES COVERED
	June 1998	Final – Oct 1995 – August 1997

4. TITLE AND SUBTITLE
An Approach to Evaluating Distance Learning Events

5. FUNDING NUMBERS
PE: 0602785A
PA: A791 TA:2233

6. AUTHOR(S) Robert A. Wisher, U.S. Army Research Institute
Christina K. Curnow, George Washington University

7. PERFORMING ORGANIZATION NAME(S) AND ADDRESS(ES)
U.S. Army Research Institute for the Behavioral and Social Sciences
ATTN: TAPC-ARI-II
5001 Eisenhower Avenue, Alexandria, VA 22333-5600

8. PERFORMING ORGANIZATION REPORT NUMBER

9. SPONSORING/MONITORING AGENCY NAME(S) AND ADDRESS(ES)
U.S. Army Research Institute for the Behavioral and Social Sciences
5001 Eisenhower Avenue, Alexandria, VA 22333-5600
Alexandria, VA 22333-5600

10. SPONSORING/MONITORING AGENCY REPORT NUMBER
Technical Report 1084

11 SUPPLEMENTARY NOTES

12a. DISTRIBUTION/AVAILABILITY STATEMENT
Approved for public release; distribution is unlimited.

12b. DISTRIBUTION CODE

13. ABSTRACT *(Maximum 200 words)*:

In accordance with an agreement with the National Guard Bureau (NGB) a simplified evaluation form was developed that assessed demographic, instructional, and technology factors for distance learning training events of short duration. The evaluation form was derived from the good practices identified in the research literature, and was used during the assessment of eight distance learning events given by the NGB nationwide. Based on returns (n=1,044) the evaluation form proved to be a reliable instrument which demonstrated adequate face validity. Based on these results, a slightly modified form was developed for use by the NGB for all short-term, distance learning events.

14. SUBJECT TERMS Distance learning, training, evaluation, Reserve Component, National Guard

15. NUMBER OF PAGES 45

16. PRICE CODE

17. SECURITY CLASSIFICATION OF REPORT	18. SECURITY CLASSIFICATION OF THIS PAGE	19. SECURITY CLASSIFICATION OF ABSTRACT	20. LIMITATION OF ABSTRACT
Unclassified	Unclassified	Unclassified	Unlimited

Technical Report 1084

An Approach to Evaluating Distance Learning Events

Robert A. Wisher
U.S. Army Research Institute

Christina K. Curnow
George Washington University
Consortium Research Fellows Program

Advanced Training Methods Research Unit
Robert J. Seidel, Chief

U.S. Army Research Institute for the Behavioral and Social Sciences
5001 Eisenhower Avenue, Alexandria, Virginia 22333-5600

June 1998

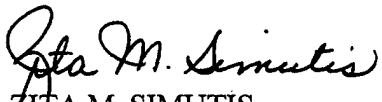
Army Project Number
2O262785A791

Education and Training Technology

FOREWORD

The U.S. Army Research Institute (ARI), through a memorandum of agreement with the National Guard Bureau (NGB) and the Army National Guard (ARNG), provides technical support to NGB initiatives on distance learning technologies. In particular, ARI was funded for evaluating distance learning events conducted by the NGB during Fiscal Years 1996 and 1997.

The present report describes the development and application of an evaluation instrument that is responsive to the NGB's interest in how well the distance learning technology worked and how well the soldiers learned. The product of this work is a streamlined, one-page evaluation sheet appropriate for evaluating any distance learning training event (one day or less) that the NGB might provide to its soldiers.



ZITA M. SIMUTIS
Technical Director

AN APPROACH TO EVALUATING DISTANCE LEARNING EVENTS

EXECUTIVE SUMMARY

Research Requirement

In 1995 the U.S. Army Research Institute (ARI) signed a Memorandum of Agreement with the National Guard Bureau to serve as the program manager for the academic component of the Army National Guard (ARNG) distance learning network. As part of that agreement, ARI agreed to conduct evaluations of various distance learning events. The research reported here concerns the development and application of a streamlined evaluation instrument appropriate for evaluations of all short-term training events conducted by the ARNG through distance learning technologies.

Procedure

The participants were n=1,044 soldiers at NGB distance learning facilities nationwide. A total of eight distance learning training events were sampled in topic areas including: Risk Management, Airborne Call for Fire, Terrorism Update, Airspace Management, Creative Problem Solving, and a Counter Drug Broadcast. A one-page evaluation instrument concerning the technology employed and the amount learned was completed by the participants. Data were analyzed on: 1) the relationship between the amount learned and previous course experience, 2) participant ratings of the technology and, 3) mode of instruction.

Findings

When three iterations of the same course were compared, evaluation results were highly reliable. Analyses revealed a greater degree of perceived learning among those participants who had not taken a previous course than those who had, also supporting the face validity of this measure. There were no significant differences of self-assessed learning among the three events for those who had taken a previous course. The similarity of responses between groups of participants over three iterations support the reliability of the evaluation measure. Participants in all eight distance learning courses were asked to rate ten aspects of the technology and mode of instruction, five of which were the same as those used in a benchmark Navy study. Ratings on these five aspects were similar between the ARNG events and the benchmark.

Utilization of Findings

This research demonstrates utility of compressed evaluation forms that are convenient to administer and effective in revealing technology and learning issues of interest to the NGB. Appendix B presents the recommended form for use by the NGB. It may be replicated for use. Appendix C is a civilian-oriented form offered for adaptation by civilian agencies.

AN APPROACH TO EVALUATING DISTANCE LEARNING EVENTS

Contents Page

Tables

Figures

Appendices

Introduction

Distance learning is becoming a growing force in the education and training industry. While businesses, the military, and educational institutions hasten to convert traditional courses, training programs, and classrooms to a distance learning format, there remains a need to verify that students are indeed acquiring the knowledge and skills being taught. Certainly, the efficacy of distance learning, when compared to traditional education and training settings, has been demonstrated hundreds of times (Russell, 1996). The studies reported, however, have been largely oriented to college courses, continuing education credits, or professional development training that occur over extended periods and which have "built-in" evaluation measures, such as a final exam. There has been little technical work in developing evaluation measures for short-term training events.

The purpose of this report is to describe an approach to creating a simplified form for evaluating the relevance and effectiveness of a distance learning event. Here, a distance learning *event* refers to a training or educational program occurring within *one day*. This is in contrast to the more frequently reported evaluations for a course or program of instruction which might last from several days to several months. For these there are usually objective, end-of-course measures that gauge achievement, or longer term assessments that measure improvements in job performance. For lengthier courses there can be more thorough evaluations such as the Course Evaluation System, a methodology used to evaluate the quality of classroom training developed by Ellis, Knirk, Taylor, and McDonald (1987). For the abbreviated training event, however, the development of objective measures of learner outcome is often not feasible as the brevity of the event precludes the justification for a lengthy evaluation. Indeed, in some cases, the time required to evaluate a training event can nearly equal the time needed to deliver the event. Furthermore, it is not always warranted to relate short training events to job performance measures.

Such is the case for a series of training events delivered on a nationwide basis to members of the Army National Guard (ARNG). This report will provide an overview of distance learning initiatives in the ARNG, review the literature on evaluations of distance learning programs, describe the rationale behind the development of an abbreviated evaluation instrument, and

present findings from the use of the instrument for numerous training events delivered to over one thousand soldiers. The strategy, then, was to develop a concise evaluation format that was both efficient in terms of administration and rich in terms of understanding training effectiveness and quality of delivery.

Distance Learning in the Army

Distance learning is not new to the Army. Indeed, the Training and Doctrine Command (TRADOC) is advocating widespread use of distance learning in the near future (TRADOC, 1996) and the ARNG has responded by developing a companion plan to connect classrooms in 54 States and Territories as one distance learning network (National Guard Bureau, 1996). In the education and training literature, distance learning has various definitions. It is generally regarded as learning that occurs in a different place from the teacher (Moore and Kearsley, 1996). Educational television is a classic example. The arrival of satellite and terrestrial networks as well as the widespread use of interactive learning techniques has propelled distance learning into the forefront of educational change. In a recent study on the use of technology for training and education in the Army, the Army Science Board recommended that the Army "continue to develop and acquire modern classroom technology, but emphasize a move toward distance learning" (Grum, et al., 1995, p.37). Indeed, universities, state and federal agencies, trade organizations, and industry now offer thousands of courses in a distance learning format. Distance learning is also of great interest to our allies, with NATO taking a keen interest in its development and implementation (Seidel and Chatelier, 1994).

Studies of distance learning in the Army have demonstrated positive results. For example, the use of the distance learning tool of asynchronous computer conferencing for the Engineering Officer Advance Course showed no difference between resident and distance learning students on objective measures of performance, but the distance learning course cost less than the resident version (Phelps, et. al., 1991). In another example with soldiers from the Reserve Component (RC), the application of two-way interactive video for training three military occupational specialties to RC soldiers demonstrated performance as high as soldiers trained in a resident mode (Bramble and Martin, 1995). In another study that examined the training effectiveness of one-way and two-way video, Lehman and Kinney (1992) report that one-way VTT performed better than two-way VTT for a course on common leader training to noncommissioned officers from the RC. In a study that measured the cost effectiveness of

audioteletraining for the Army National Guard, favorable results of equal effectiveness at a lower cost were reported for the training of unit clerks (Wisher, Priest & Glover, 1997). Finally, in a bibliographic review reported by Howard (1997), 106 articles and books were identified on distance learning topics relevant to the Army. Only one reported a negative effect of distance learning.

On a broader level, distance learning is viewed by the Army as the "*delivery of standardized individual, collective, and self-development training to soldiers and units at the right place and right time through the application of multiple means and technology*" (TRADOC, 1996). Because of the numerous training technologies this definition implies, it is useful to regard distance learning from a "toolbox" point of view, where different instructional tools that can be selected and applied to satisfy a training requirement. The choice of tool (video teletraining, computer based training, print material, etc.) depends on the type of task being taught, the purpose of the training, the size of the training audience, cost, and the availability of facilities, equipment, and courseware. Such an approach allows for training media to be allocated in appropriate combinations for a course. The mixing of media for distance learning has a long tradition. For example, the British Open University makes significant use of media combinations for most of its courses (Brown, Nathenson, & Kirkup, 1982).

The emergence of the Internet as a resource for instructional delivery furthers the application of multiple media, as streaming multimedia technologies can offer audio, video, and three-dimensional animation for use by the creative instructional designer. It is hard to forecast what new or improved tools will emerge years from now for use in instructional technology. Nevertheless, there remains a need to verify that students continue to learn in this evolutionary environment.

Background

Evaluation is concerned with judging the worth of a program: is the program effective?, does it save money?, can we improve its capability based on feedback? Depending on the level of evaluation, there is generally a tradeoff between the resources for evaluation and the decisiveness of results. A measure of student reaction to training, for example, is less compelling than a measure of the degree to which that training transfers to job performance. However, depending on the resources supporting the evaluation, the costs and time to develop and

implement more thorough measures of learning might be impractical. The tradeoff between resources and decisiveness of results will be discussed later.

The evaluation of technology-based programs in education and training has grown commensurately with the increased efforts to implement technology in the classroom and other learning settings. Indeed, Baker and O'Neil (1994) contend that "technology assessment studies attempt to make predictions about use to a broad class of potential users and settings, and to inform policymakers' future technology investment strategies" (p. 6). Baker and O'Neil also identify a technology assessment strategy in which a series of individual evaluation studies, either in serial or parallel designs and conceived as natural experiments, combine to provide coordinated information to the decision makers. In the context of distance learning, researchers such as Clark (1994) call for multilevel evaluation plans that have measures of both participant reactions and the achievement of program objectives. Both the multilevel plan and parallel design concept were incorporated in the present approach to evaluating distance learning events.

Early evaluation studies in distance learning were mostly descriptive case studies that focused on learner satisfaction, and were often anecdotal (OTA, 1989). They were often conducted as an after thought and relied on reaction questionnaires which were often unreliable or not representative of the students involved. More recently, there have been discussions regarding what methods are best to use for evaluating distance learning programs (Harrison, et. al., 1991). This study proposed a more comprehensive examination of distance learning as a complex system involving an interaction of organizational, administrative, instructional, and technological components. In their review of the evaluation literature, Harrison et. al. (1991) identified three unique components of distance learning that consistently emerged: instruction, management, and logistics (i.e., technology factors, technical qualities, environment, on-time delivery). From this point of view, the National Guard Bureau was primarily interested in the instructional and technology factors of effectiveness.

Evaluation Measures of Instruction

As with evaluations of other modes of training in organizations, the instructional effectiveness of a distance-learning training event can be measured in terms of student reaction, learning, behavioral criteria, or results criteria (Wexley and Latham, 1981). Each measure is used to assess different aspects of the value of an event. Some are objective and some are

subjective, and each has advantages and disadvantages in terms of preparation of instruments, administration time, and ease of interpretation. The four measures will be briefly reviewed.

Student Reaction Measures. An evaluation based on reactions measures how favorably the participants respond to a training event, including its contents and delivery, relevance to their job, the training technologies used, the quality of audio and video, and the overall training environment. Favorable reactions to training events as measured by this criterion, however, do not guarantee that learning has taken place. Learners can mistake good presentation style for good instruction (Ghodsian, Bjork, & Benjamin, 1997). For example, an event may have had both an entertaining instructor and excellent technical qualities, and, consequently, student reactions were highly favorable. However, if learning had been measured more directly, there may have been no significant pre to post difference in knowledge.

Although favorable reaction measures alone do not warrant that learning really occurred, they are useful to collect for several reasons. First, positive reactions help gain or maintain organizational support for training events. If there were negative reactions to events, such as subjective participant ratings that nothing was learned or that the participants perceived the program as being a waste of time, support by the organization to continue such training might quickly erode. Positive reactions do not guarantee learning, but negative reactions can lead to problems of future resourcing and support.

The second benefit is reaction measures can serve as a source of immediate feedback to the training providers, including instructors, production staff, and training event organizers. Such data can prove helpful in planning and designing future training events. For example, low ratings for responsiveness to questions may inform the instructor to include more time for a question and answer session. Problems with technical qualities, such as poor audio quality, at specific sites can identify problem areas for correction. Also, specific questions can be formulated to address potential concerns, such as was the broadcast time convenient if the training spanned numerous time zones. Furthermore, repetitive measures of the same training delivered over a period of time can demonstrate progress in improving the instructional quality and technical delivery.

A third benefit, not unique to reaction measures, is the insight that can be gained from subgroup analysis: sample is divided into two or more subgroups, and then comparisons of reactions across these groups are made. This process is referred to as cross tabulation, or

contingency table analysis (Babbie, 1990). Such an analysis can elucidate the varying impact of the training across subgroups. For example, let us assume that half of the participants (Subgroup A) in a training event reported prior training on the event's topic. The other half (Subgroup B) reported no prior training. If the cross tabulation reveals that Subgroup A reports little knowledge gain on the topic, in contrast to Subgroup B reports a substantial gain in knowledge, it is clear that the level of training was not appropriate for the entire audience. In this case, event organizers might base participation in later iterations on a selection factor that includes only those with no prior training on the topic. A more advanced event might then be scheduled for those with more experience. Without such an analysis, this useful finding could have been obscured in by an average rating of "some" knowledge gain. There are many subgroups that can be formed from variables, or combinations of variables, within the reaction measure instrument. (It should be noted that the subgroup analysis assumes that there is statistical independence between groups.)

Self-Assessment. Another type of reaction measure, obviously subjective, is the self-assessment variable. This measure requires that individual participants evaluate themselves on various dimensions of the training, such as how well they paid attention to the instructor or how much more they learned compared to what they already knew about the topic. The self-assessment variable may be viewed as a bridge between a reaction measure and a learning measure, as it assesses learning directly, but in a subjective way.

Self-assessment is obviously an easy technique to judge a student's knowledge or skill. Students are simply asked how much they learned, how high they could score on a hypothetical test, or how well they could perform a task. The validity of these judgments is debatable. Regardless of the context of the self-assessment measure, most of the research in this area has identified significant correlations between self-assessments and performance measures (either knowledge and ability tests or supervisor assessment). For example, positive results have been reported for self-assessment of: job performance and supervisor assessments (i.e. Meyer, 1980; Harris & Shaubroeck, 1988); typing test scores and actual typing scores (Levine, Flory, & Ash, 1977); performance and objective performance measures (Farh & Dobbins, 1989); and abilities and ability tests (DeNisi & Shaw, 1977). Schendel, Morey, Grainer, and Hall (1983) found significant correlations between self-assessment of M16A1 rifle qualification scores and actual performance.

Another general finding in the self-assessment literature, one that has been the basis of criticism, is that self-assessments tend to consistently overestimate actual performance (i.e. Schendel et. al., 1983; Anderson, Warner, and Spencer, 1984). The latter study found that this bias to overestimate could be corrected for by rescaling based on information from "bogus" questionnaire items. The overestimation bias, however, is relatively constant, so that there is a consistent difference between the self-assessment scores and the performance scores of an individual.

In research most relevant to the current report, a secondary analysis of performance data collected during an Army mobilization training exercise is reported in Curnow (1998, in preparation). This analysis supports the idea that in military training environments, self-assessment is a fairly accurate indicator that does not necessarily overstate one's abilities. A soldier is accustomed to being regularly evaluated on a strict Go / No Go basis throughout a career. This constant feedback probably contributes to a metacognition of individual learning.

Learning Measures. Learning measures offer a more objective assessment of the knowledge and skills acquired during a training program. Knowledge refers to the facts, principles, rules, and procedures that were taught, and can generally be measured through paper-and-pencil tests. Skills generally refer to the application, or transfer, of what was acquired in the classroom to a time and event dependent environment such as the workplace. Skills are generally measured through hands-on performance tests or situational exercises in which speed as well as accuracy of responses are of principal concern.

Although learning criteria are a stronger, more decisive measure than reaction criteria, a drawback is the time to develop test instruments and the costs of their administration – the evaluation resources. Written and/or performance tests specific to the knowledge taught need to be developed, pre-tested, and then administered under a controlled test environment. Ideally the reliability and validity of the tests should be determined prior to the training event. For many traditional courses taught through distance learning, such tests are often available in the form of a final test for academic courses or practical exercises, end-of-course tests, and hands-on performance tests for training a military occupational specialty (Wisher et. al., 1997). For a short duration training event, however, a learning measure usually requires additional development effort which might be impractical for a short-term event. The use of self-assessment as a

surrogate learning measure, particularly in a military setting, might remedy this issue. This is a tradeoff that must be judged on a case by case basis.

Behavioral Criteria Measures. Behavioral criteria are concerned with the follow-up performance of the participant in another environment, such as a job setting. The issue is whether what was learned in the training event transferred to the workplace. For example, consider a case where every employee in an organization participates in a training event on "Ethics in the Workplace." The event may have received a favorable reaction from participants. As measured by learning criteria, the event may have increased participants' knowledge of the topic. However, if there is no comparative reduction in the target behavior over a pre-determined period, that is ethical violations in the organization remain at the previously measured level (which was unacceptably high), then the training event was unsuccessful. Behavioral measures require a period of time, weeks or months, before the effectiveness of training can be judged, and it is helpful to include a control group. The resources for a behavioral measure, especially the time factor, can be high.

Results Criteria Measures. Results criteria are similar to behavioral criteria in that they are also concerned with the performance of an individual, but usually in the context of a productivity gain. A classic example is a sales training event. As in the previous example, a sales training event may have had favorable reactions from the participants, and measures of their learning were positive. But if there was no comparative gain in sales within the region over a pre-determined period, then the training event was unsuccessful on the basis of a results criterion. This measure also requires data collection over an extended period. Both the behavioral criteria and results criteria measures are more decisive than either reaction or learning measures due to their "bottom line" orientation.

Resource Tradeoffs. The choice of which measure to use depends on the specifics of the training as well as some practical considerations, such as time and costs. The more time and dollars available, the greater the resources to develop a knowledge test and track changes in behavior or results of performance over an extensive period of time. On the other hand, if results are sought quickly and resources are limited for the development, administration, and analysis of data, a reaction measure may be the only option. A notional relationship between the four evaluation measures described above ranked by resource requirements and the decisiveness of the results derived from the measure is depicted in Figure 1.

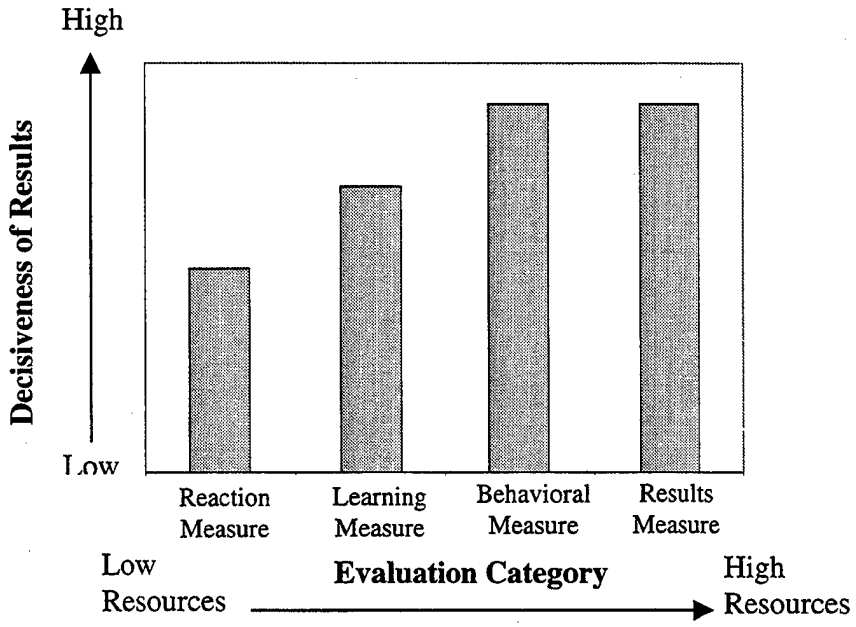


Figure 1. Notional tradeoff between evaluation resources and decisiveness of results.

In the context of distance learning training events for the Army National Guard (ARNG), the reaction and learning measures are most relevant due to the need for quick feedback to event organizers and sponsors as well as to monitor whether the training objectives of the events are being met. One of the practical considerations was the anticipated effort required for the development of a knowledge or performance test for the learning measure. In view of the limited resources for instrument development and the range of anticipated training events, a strategy to have a self-assessment scale serve as the learning measure was selected. This set the stage for a compressed evaluation form to be created and the subsequent evaluation time to be brief, since a lengthy knowledge test would not be included. More importantly, it provided an acceptable source of data for the types of analyses relevant to the ARNG's interests, mainly the instructional and technology factors. Other strategies in the development of the evaluation form considered factors that would lead to a high response rate, as discussed below.

Development of the Compressed Form

Factors

Evaluators always seek a high rate of return in order to reduce errors in measurement due

to of a small sample. This is especially true for evaluations conducted at remote sites. The evaluation form under consideration was to be administered to participants who had just received training at dozens of remote sites. Usually, the training was of short duration, sometimes one or two hours, but not more than a full work day. In view of the brevity of the training, a goal was to develop a form that would be both quick for the participants to fill out and responsive to the interests of the National Guard Bureau: how well did the technology work and did the soldiers learn from the training experience? In consideration of these issues, a review of the literature was conducted. Four topics were identified as being conducive to a quick response and a high rate of return: length, saliency, confidentiality, and ease of return.

Length of Instrument. One method to increase the response rate is to shorten the length of an evaluation form. The research dealing with questionnaire length and response rate, however, has been ambiguous. Kanuk and Berenson (1975) identified seven studies that examined questionnaire length and found that only one reported a statistical relationship between questionnaire length and response rate. The research on this topic has been limited mostly to the areas of marketing and consumer satisfaction. The research most relevant to the present report was a study reported by Roszkowski and Bean (1990), in which questionnaire length was controlled in a mail survey to over 8,000 persons taking a distance education course in insurance and financial planning. In 14 replications, persons were randomly assigned to receive a long or short course evaluation questionnaire. Response rate for the short form was significantly higher in all 14 replications, averaging a 28% advantage. The authors conclude that, in the context of distance education, shortening the questionnaire did improve the rate of return.

Saliency of Instrument. Saliency refers to the importance or relevance of the questionnaire content to the respondent (Altschuld and Lower, 1984). Based on three studies, Baumgartner and Heberlein (1984) concluded that salient surveys result in higher response rates. They found that surveys judged to be highly salient obtained a 77% response rate, while those judged not to be salient obtained a 42% response rate. Baumgartner and Heberlein's (1984) results have been duplicated by Goyder (1982) and also by Eichner and Habermehl (1981). For a distance learning evaluation, questions related to the course, instructor, and the technology employed should be perceived as highly salient.

Confidentiality. Another factor found to effect the return rate of questionnaires is guaranteed confidentiality (Altschuld and Lower, 1984). Confidentiality refers to the promise

that data that could identify an individual respondent's answers will not be released. This confidentiality can be expressed on the evaluation form through a statement that the information collected is in accordance with the Privacy Act of 1974 and that "Information on individuals is confidential and will not be released to anyone." Related to the confidentiality factor is anonymity, which refers to the lack of information that can uniquely link a specific set of answers to a specific respondent. Not asking for a respondent's name communicates an intention for anonymity, but it does not guarantee it, since demographic responses coupled with location information might identify an individual. For example, in the context of the National Guard, if there is only one respondent of the rank of Colonel, then data from that Colonel should not be reported separately in a cross tabulation. Rather, the data should be netted with other ranks, such as with Lieutenant Colonels and Majors and reported under "field grade officers." A combination of the Privacy Act statement along with not asking for names is likely to lead to improved return rates.

Convenience for Return. Ease in returning evaluations also increases response rates. In a meta-analysis of response behavior, Yammarino, Skinner, and Childers (1991) found significant, positive correlations between the inclusion of a pre-addressed return envelope and response rate. In addition, they found that for institutional (rather than consumer) populations, stamped or metered postage increased response rates. Both a return envelope and prepaid postage make returning an evaluation more convenient, hence increasing the response rates.

Examples from the Literature

The recommendations from the literature on length, saliency, confidentiality and anonymity, and return were factored into the design of the training event evaluation form. Additional suggestions were gained from other evaluators in the distance learning field. A review of instruments from educational institutions and government agencies, most published in the open literature but some made available through personal communications, resulted in the creation of a set of comparative variables: number of pages for an instrument, number of assessment categories (such as demographic, instructor, facilities etc.), number of questions within each category, and a description of the scales used (Likert, yes/no, fill in, etc.). The results of this analysis are displayed in Table 1.

Table 1

Characteristics of Distance Learning Evaluation Instruments

Reference	Page Count	Categories (Items)	Response Scale
NGB Student Evaluation Form Wisher and Curnow (Present Report)	1	Course (5) Instructor (3) Technology (8) Motivation (1) Demographics (5) 5 categories, 22 items	5-pt Likert (11) Yes-No (2) Categorical (7) Miscellaneous (2)
Jones (1992) Interactive Telecommunications System (IITS) Students' evaluation.	5	Course (9) Instructor (23) Technology (18) Motivation (10) Demographics (7) Comparisons (5) Miscellaneous (5) 7 categories, 77 items	5-pt Likert items (65) Yes-No (5) Miscellaneous (7)
Wetzel (1995) Navigation refresher course delivered by videoteletraining. and Wetzel, Radtke, Parchman, Seymour (1996) Fiber optic cable repair course by videoteletraining.	4	Instructor (6) Technology (8) Interaction/Participation (4) Training Aids (9) Learning & Classroom (7) Overall (4) Background (3) Miscellaneous (3) 8 categories, 44 items	5-Point Likert (35) Categorical (60) Yes-No (1) Fill In the Blank (3)
National Enforcement Training Institute. Environmental Protection Agency (personal communication)	2 & comment page	Course Objectives (6) Facilitator (3) Technology (6) Presentations (5) Exercises (4) 5 categories, 24 items	5 pt. Agree-Disagree
Simpson, Wetzel & Pugh (1995) Division officer Navy leadership. and Wetzel, Simpson, & Seymour (1995) Petty officer Navy leadership training.	2	Instructor (9) Interaction/Participation (4) Training Aids (3) Students/Teams (4) Learning Activities (4) Overall (comparisons) (3) Miscellaneous (3) 7 categories, 30 items	5-point scale (27) Fill in the blank (3)

Kabat & Freidel (1990) Comprehensive evaluation model for assessing the effectiveness of a two-way interactive distance learning system.	2	Course (2) Instructor (6) Technology (8) Demographics (3)(3) Miscellaneous (5) 5 categories, 27 items	5-point Likert (16) Yes-No (3) Categorical (3) Fill in the Blank (5)
Student Critique Video Teletraining (VTT) (personal communication)	2	Course (11) Instructor (8) Site Monitor (6) Classroom (12) Media (5) Testing (7) 6 categories, 49 items	9-Point Likert w/ not applicable
Biner, Dean, & Mellinger, A. (1994) Distance learner satisfaction with televised college-level courses.	not specified (1 or 2)	Course Mgt./Coordination (11) Instructor/instruction (14) Technology (8) 3 categories, 33 items	5-point
U.S. Army Logistics Center. (personal communication)	1 & comment page	Course (6) Instructor (3) Technology (4) Miscellaneous (4) 4 categories, 17 items	9-point scale
Mackin, D. & Hoffman (1996) Focus groups over interactive television.	1	Course (6) Demographics (6) 2 categories, 12 items	Demographics 4-point Comments
Jegede, Gooley, & Towers, (1996) An evaluation audiographic conferencing professional development programs.	not specified	Course (36) Instructor (7) Interaction/Participation (8) Training Aids (4) 4 categories, 55 items	5-point Likert w/ not applicable

The NGB Student Evaluation Form, developed in the present report and listed as the first entry in Table 1, was compressed into a single page based on the factors described earlier. The form is presented in Appendix A. There are five assessment categories with a total of 22 items. The respondent completion time is about 3 minutes. Of the 22 items, 18 are related to the course, technology, and demographic factors. This selection was based on the National Guard Bureau's basic interest of "did the technology work and did soldier's learn?" The demographic variables were needed to detail a more refined look at learning patterns that might emerge from a cross tabulation of the data. The goals of a short form with a confidentiality statement, salient to-the-point items, and a pre-paid return envelope provided to each site were accomplished.

Test Administration of Compressed Form

Eight distance learning events were sampled for inclusion in the analysis. These events represented the primary distance learning events available on a nation-wide basis for members of

the Army National Guard during the test administration timeframe. All training was delivered through a multi-point, one-way video two-way audio connection through satellite links. Table 2 provides a summary accounting of the training events, showing the date of each event, the reported attendance, the number of surveys that were returned, and the return rates. The events can be categorized broadly as military related or civilian related. The military events addressed specific training requirements related to military jobs and operations. The civilian events were general in nature, addressing areas that would be of interest to both civilian and military organizations.

Table 2. Training events sampled for the analysis

Training Event	Date	Origination Site	Number of Remote Sites
Military Training Events			
Risk Management	7, 8 Jun 1996	Fort Lee, Virginia	37
Risk Management	5, 6 Oct 1996	Ft. Rucker, Alabama	13
Risk Management	4, 5 Apr 1997	Ft. Rucker, Alabama	16
Airborne Call for Fire	13 Jul 1996	Arizona State University	4
Terrorism Update	19 Dec 1996	Emmitsburg, Maryland	54
Airspace Management	17 May 1997	Arizona State University	94
Civilian Training			
Creative Problem Solving	20 Aug 1996	Dallas, Texas	10
Say it Straight	25 Mar 1997	St. Petersburg, Florida	193

Military Training Events

The Risk Management event (2 hours) was taught during three periods (by the same instructor) over the course of ten months. Risk management refers to the governance of safety practices during training and field operations. The training originated from the Satellite Education Network facility at Fort Lee, Virginia, for the first event and from the Army National Guard Media Center at Fort Rucker, Alabama, for the other two iterations. The broadcast was designed to cover risk management and risk assessment in support of mission requirements. Training topics included understanding the risk management process, the rules of risk management, and the integration of these considerations into mission planning. The target audience was supervisors at all levels, commanders at the battalion and company level, mission planners, and appointed safety and health personnel.

The Airborne Call for Fire event (6 hours) was designed to familiarize helicopter-borne, forward observers from the artillery branch with the techniques and capabilities of calling for

artillery fires from mobile helicopter platforms. It was sponsored by the Western Area Army National Guard Aviation Training Site (WAATS) and originated from Arizona State University.

The Terrorism Update event (4 hours) originated from a Federal Emergency Management Agency (FEMA) in Emmitsburg, Maryland, and was marketed in a cooperative effort by the Army National Guard and FEMA to both an internal ARNG audience and to external law enforcement, fire, medical responder, and state/county emergency preparedness audiences. This was an attempt to demonstrate the "community use" potential of an ARNG Distance Learning Network. The program content addressed new anti-terrorism legislation and funding initiatives as well as lessons learned from the 1996 Summer Olympics and the 1996 Democratic National Convention.

The Airspace Management event (2-hours), sponsored by the WAATS and originating from Arizona State University, was designed to cover various facets of new air space requirements instituted by the Federal Aviation Administration.

Civilian Training Events[1]

The Creative Problem Solving event was geared towards today's government professional who needs to be more creative to stand out, succeed or even survive in an every changing organization. It was sponsored by the U.S. Department of Agriculture Graduate School and originated from Dallas, Texas. Participants were trained on how to improve problem solving abilities in the workplace, tap into their own creative potential, conceive new strategies and new ideas, and figure unconventional ways to do their jobs better.

The Counter Drug broadcast event, known as "Say it Straight," concerned the prevention of drug use, with a focus on the medical consequences of drug use and efforts to legalize drugs. It was sponsored by the Multijurisdictional Counter Drug Force Training Academy and originated from St. Petersburg, Florida. It had a projected audience of over 8,000. Because of this large number, a stratified, random sample of sites was taken. The sites were stratified based on geographic region and estimated number of attendees, so that the final sample contained sites

[1] Due to restrictions by the Office of Management and Budget on surveying private citizens, the data to be reported are from only members of the Army National Guard who participated in the training events. Some civilians may have participated in the training event, but they were not included in the analysis.

from each of seven regions, and participants in small (<25), medium(26-74), and large (74+) audiences.

Administration Procedure

An address list of participating sites was compiled for each training event. Each site was sent a package containing a cover letter, course materials, instructions to the site facilitator, evaluation forms, and a postage paid, pre-addressed return envelope. Site facilitators were asked to distribute the evaluation forms immediately after the event was completed and have the participants complete the form before leaving. The forms took no more than a few minutes to complete. The facilitator then mailed the return envelope. For each event, there were a few sites that either canceled at the last minute or experienced significant technical difficulties, such as no signal. Those sites were not included in the response rates. When the competed evaluations were received, the data was compiled using an optical scanner and then converted to a format compatible with statistical analysis software.

Descriptives of Sample

The first three items on the evaluation form were demographic questions concerning rank, education, and age. As can be seen in Table 4, the rank variable consisted of six categories. The largest subgroup of respondents were non commissioned officers in the E5 to E9 paygrade (32%). The second largest group of respondents were Majors or above (23%), with 37% of the entire sample being Second Lieutenants or higher.

Table 3

Rank of Respondents

Rank	n	Percent
Private to Specialist (E1 to E4)	24	2%
Non Commissioned Officer (E5 to E9)	330	32%
Warrant Officer	159	15%
Company Grade Officer (O1-O3)	144	14%
Field Grade Officer or higher (O4+)	236	23%
Army Civilian Employee	140	14%

The education question asked respondents to report their highest level of education, 6% responded high school diploma, 27% reported some college, but no degree, 13% reported having an associates degree, 39% reported having a bachelor's degree, and 14% reported having an advanced degree. Over half of the sample had a bachelor's degree or higher.

The average age for the respondents was 41 years with a range from 20 to 59. This was of interest because there was a concern within the National Guard that the "older" soldier might not be as responsive to technological change as the younger soldiers. (This was not the case. The age variable was not correlated (higher than $r = \pm 0.1$) to any of the course and technology ratings.)

Findings

Return Rates

For each event, the large majority of the evaluation forms were returned within five days of the event. Site facilitators returned attendance lists to an event coordinator, allowing return rates to be computed. Table 4 shows the return rates for each event. The percentage of returned evaluation forms ranged from 30% to 97%, with a mean return rate of 74%. (The Counter Drug event had a very large audience including cable viewers, so a list of actual attendees was not available. Therefore, a return rate could not be computed, and this event does not appear in Table 4.)

Table 4

Return Rates

Training Event	Reported Attendance	Eval. Forms Returned	Return Rate
Military Training Events			
Risk Management	406	227	56%
Risk Management	101	88	87%
Risk Management	261	234	90%
Airborne Call for Fire	90	87	97%
Terrorism Update	352	297	84%
Airspace Management	26	18	69%
Civilian Training			
Creative Problem Solving	70	21	30%
Say it Straight	97	72	--
	1306	972	74%

Four questions addressed the respondent's previous experience with interactive training events. As noted in Table 5, approximately one-third of the respondents reported attending at least one other interactive TV event either for Guard/Reserve training, civilian job, or education or occupational training, only 10% reported attending at least one interactive event for other

interests (e.g. hobbies). For a separate question, 58% of the respondents reported previously taking a course in the topic of the training that they were attending.

Table 5

Percentage of Respondents Who Reported Attending Previous Interactive TV Training Events

Type of Training	Percent of Respondents Who Reported Attending At Least One other Interactive TV event
Guard/Reserve	38%
Civilian Job	30%
Educational/Occupational	32%
Other (e.g. hobbies)	10%

Nine questions addressed technology factors, the course, the instructor, and the learning environment. Each question asked respondents to rate an aspect of the event using a five-point scale from "poor" (1) to "excellent" (5). The means and standard deviations for each appear in Table 6. Also presented are means from substantially similar questions reported in studies of distance learning conducted by the Navy (Wetzel et. al., 1996; Simpson et. al., 1995). These provide a benchmark comparison for the ratings obtained in the present study. By and large the students reacted favorably to the experience in agreement with the Navy study. The lowest mean ratings occurred for "opportunity to ask questions" and "quality of audio" and the highest ratings were for "location of video" and "quality of video." However, it is the variability among questions that can lead to useful interpretations. For example, "quality of audio" had a relatively high standard deviation of 1.21, which obviously indicates some low ratings were present. When the audio scores are examined by remote site for each event it becomes clear that some sites probably had serious problems with their audio equipment. For example, in the first Risk Management course, a one way ANOVA showed significant differences for "quality of audio" between sites ($F(20, 196)=5.97$, $p<.001$), more specifically a post hoc Sheffe test showed significant differences between two particular sites with means of 4.63 and 2.64 ($p=.016$). Feedback to the technical staff on this type of disparity is essential.

Table 6

Means and Standard Deviations for Student Ratings of Events.

Question	Navy Benchmark	ARNG Events		
		M	SD	n
Location of the video screen	4.6	4.4	.81	1031
Quality of audio	4.1	3.8	1.21	1023
Quality of video	4.5	4.2	.96	1025
Instructor Effectiveness	---	3.9	.91	960
Opportunity to ask questions	---	3.7	1.22	903
Responsiveness to student questions	4.5	3.9	1.00	849
Relevance of course to guard duties	4.3	4.1	.96	938
Overall learning environment	---	4.0	.99	1005
Overall effectiveness of instruction	---	3.8	.98	1000

The final question on the evaluation referred to travel time. This question was included since one of the goals of the National Guard Bureau is to have distance learning facilities within one hour of every soldier's home. In this sample, 81% of the respondents traveled for less than one hour to attend the event. The sample, however, was not random so this figure does not necessarily represent the current proportion of soldiers within one hour of a distance learning facility nationwide.

Self-Assessment Measure

A self-assessment measure was used in the evaluation form as a surrogate for a learning measure. As previously discussed, the rationale for using self-assessment of the amount learned on a topic must be judged on a case-by-case basis. The form of self-assessment in the current report stems from the question "Compared to what you already knew about '*course topic,*' how much more did you learn in this training event?" Since resources for the evaluation were limited and the variety of training events was considerable, the development of more decisive measures of learning was not practical, nor were they already available from end-of-course tests. In combination with another key measure, namely previous training on a topic, the test properties of the self-assessment variable require further examination due to their overall importance in interpreting the results

Psychometric Properties

The two common psychometric properties of concern to evaluators are the validity and reliability of a measure. First, let us examine the question of validity of the self-assessment

measure. Face validity refers to whether a measure appears to make sense. In the case of the Risk Management courses, face validity was examined by comparing perceived learning between subgroups who either had or did not have a previous course on the topic, with the expectation that those who had a previous course would learn less. Analyses suggested significant differences in self-assessed learning between those who had taken a previous course and those who had not ($F(1,476)= 28.16$, $p<.001$). This finding supports the face validity of this measure, as it would make sense that those who had no previous training in a topic would likely learn more than those who had previous training. This trend was true for all of the military training events.

Reliability refers to either the consistency of a measure from one use to the next (external reliability) or the consistency of items within a measure (internal reliability). When repetitive measures of the same treatment give similar results, the measurement is said to be reliable (externally). Traditional measures of reliability, such as the alpha coefficient which measures the consistency of scale items and test-retest reliability which measures the consistency of one test administered to the same person at two separate points in time, were not an appropriate approach to measuring reliability for this instrument. The apparent consistency in the cross-tabulation of amount learned and previous course experience for the repetitive measures of three Risk Management courses was further examined to gauge the reliability. The previous course experience variable was cross tabulated by the self-assessed learning variable, and a 2 X 3 ANOVA was conducted, with two levels of previous course experience and three levels of learning. This analysis revealed no significant mean differences of self-assessed learning among the three events ($F(2, 476)= .38$, $p=.684$). The ANOVA results indicate that the amount of perceived learning was consistent across all three events. Furthermore, there were consistent differences in perceived learning between those who had and had not taken a previous course in Risk Management. This consistency of means supports the reliability of this measure.

The consistent differences between the two levels of previous course experience showed that those with no previous course always reported learning more. This offers a convenient basis on which to judge prior knowledge, allowing for adjustments to the program as necessary, such as creating an introductory and advanced version of a course.

Interesting Patterns from Cross Tabulations

Overall, 58% of the respondents reported previous training in the topic being covered. Of particular interest as the interrelationship between two variables, the amount of perceived

learning and whether or not the respondent had previous training on the topic. This interrelationship can be examined by using cross tabulation, resulting in the graph seen in Figure 2. Note that on the five point scale, ratings of 1 or 2 were netted as "little" was learned, a rating of 3 was interpreted as "some" was learned, and a rating of 4 or 5 was netted as a "lot" was learned. In all cases, the ratings were made relative to whether the respondent had previous training on the topic.

The graph shows a pattern in which those who had not taken a previous course in the topic area reported greater amounts of learning. This is evidenced in the following way. The left panel of the figure depicts those respondents who reported no previous course. More than half of those with no previous course reported learning "a lot." The right panel of the figure charts the trend for those who had taken a previous course in the topic area. The majority of these respondents reported learning at least some "some" or a "lot" more than they already knew. However, only 32% reported learning "a lot" compared to 62% in the no previous course group. While the majority of respondents reported at least "some" learning, it appears that the courses were generally geared to those individuals who had no previous training in the topic area. Such a pattern can quickly be discerned from the monotonically increasing rate in the left panel compared to the sharp drop off in the right panel of Figure 2. In comparison, many more of the respondents (26% vs. 12%) who had previous training learned "little," indicating that a more advanced course may have been more appropriate for them.

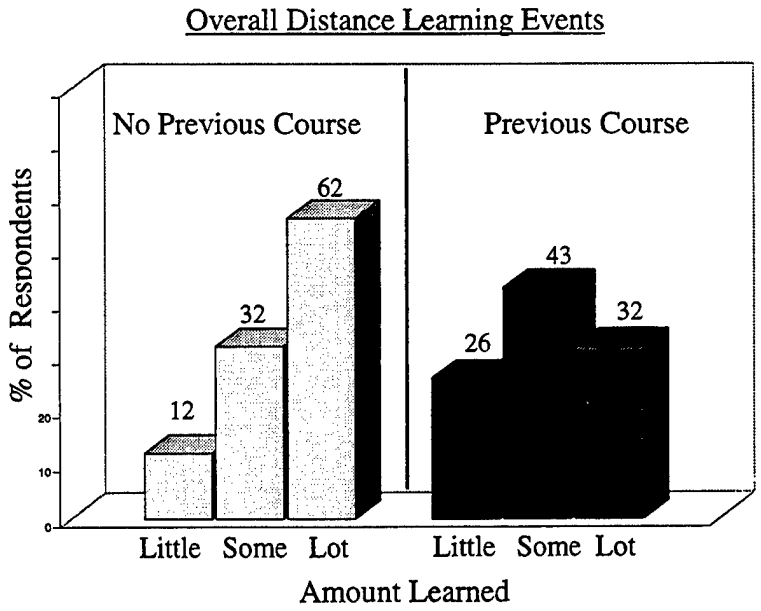
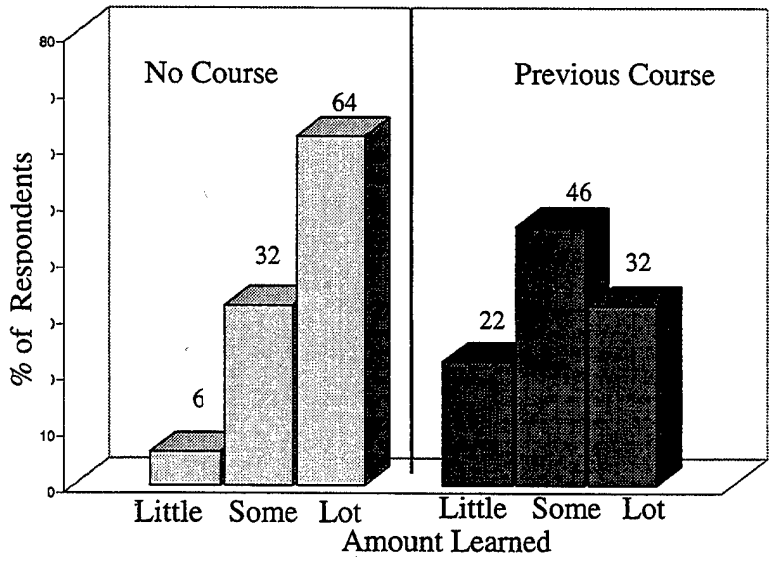
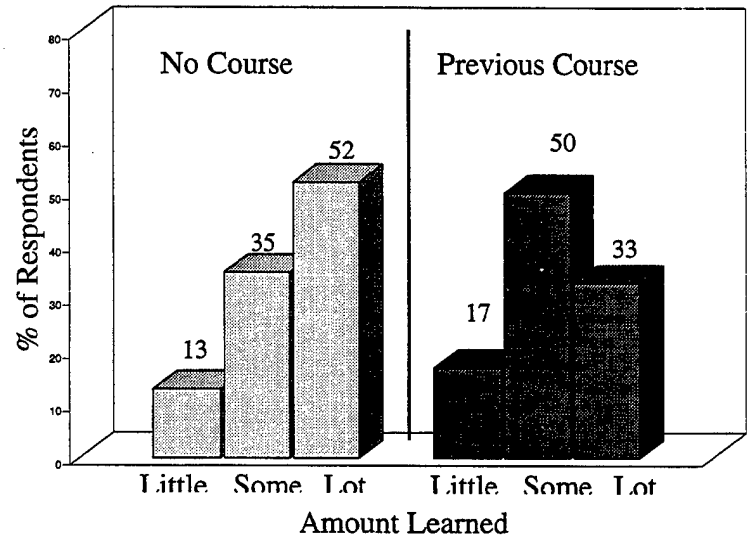


Figure 2. Amount learned cross-tabulated with previous course for overall sample.

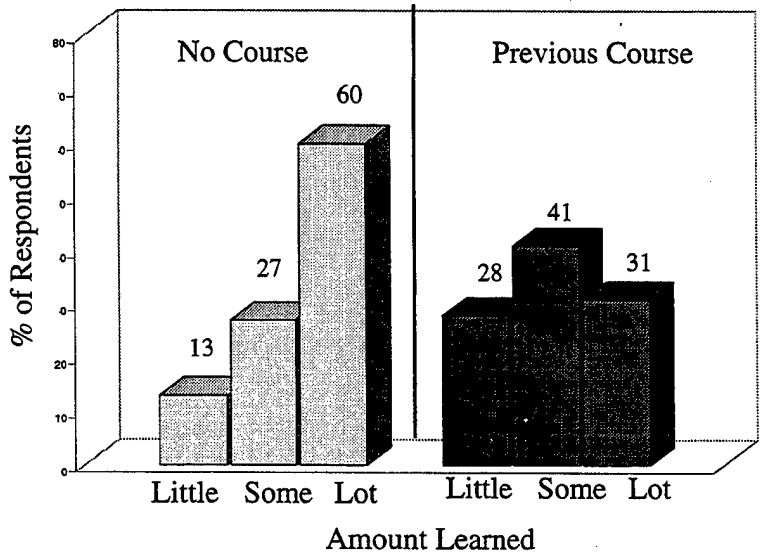
In order to explore more thoroughly the relationship between previous training and amount learned, responses were segmented into subgroups by course. As the Risk Management course was evaluated on three separate occasions, this presented the best opportunity to explore the consistency of the relationship between these two variables. As can be seen in Figure 3, the patterns between the three courses were quite similar. Those with no previous course tended to report a greater amount of learning than those with a previous course. For all three events, the greatest percentage of respondents reported learning a "lot" for the group with no previous course, while the greatest percentage of respondents reported "some" learning for the group with a previous course in Risk Management. These results are quite consistent, and could be used as feedback by the instructor to either alter the course materials, or to notify potential event participants of the level of knowledge for which the course is best suited. (The breakouts for the other military events and other civilian events are provided in Figure 4 and 5 at the end of this report.)



Risk Management (I) (n=202)



Risk Management (II) (n=88)



Risk Management (III) (n=234)

Figure 3. Subjective learning trends in three iterations of the same course.

Another variable of interest in this evaluation was the travel time to the distance learning site. The National Guard Bureau is interested, as a matter of policy, in having a distance learning facility within a one hour drive of each soldier's home. Tracking this variable, then, is a useful way to measure the trend in reducing travel time over a period of years. The travel times for this sample, listed in Table 7, may serve as a 1996-1997 baseline for the trend, keeping in mind that it is not necessarily a random sample but rather an accounting of those who attended.

Table 7

Student Travel Time (n=1,040)

Travel Time	Percent
1-29 minutes	68%
30-59 minutes	13%
1 to 1.5 hours	8%
1.5 hours or more	11%

When the travel time variable was separated into two categories (less than one hour and one hour or more) there is an interesting difference within the course ratings. As can be seen in Table 8, those who traveled farther rated certain technical aspects of the course (quality of audio and overall learning environment) significantly lower than those who did not travel as far. One explanation might be that students who traveled for an hour or more to a distance learning facility had higher expectations about the technical quality of the training event. They may have been a bit disappointed in the quality given their greater effort to get to the site. It is interesting to note that the ratings for the amount learned did not seem to be affected by travel time.

Table 8

Mean course ratings by time traveled.

	Travel Time		
Rating	Less than one hour	One hour or more	t-test
Quality of Audio	3.9	3.5	4.5*
Overall Learning Environment	4.1	3.8	4.1*
Learning Assessment	3.3	3.3	-0.8^{ns}

*significant at p<.01

^{ns} Not Significant

The findings and analyses reported above demonstrate the wealth of informative data that can be obtained from a simple, one-page evaluation form. There are many other cross-

tabulations and analyses that could be conducted, depending on the interests of the organizations, to pinpoint technical shortfalls, course effectiveness, learning, and certain policy issues.

Discussion

A useful strategy for evaluating single-day distance learning events is the use of a compressed, one-page evaluation form that is designed with considerations of confidentiality, saliency, convenience of return, and length. The research literature has demonstrated these factors to be of value in obtaining higher return rates. Another aspect of the literature that was useful was the self-assessment variable for evaluating learning. Even though this variable has had some controversy in the past, its use in a military training setting (especially when coupled with anonymity) appears acceptable in this situation, as evidenced by both the face validity and external reliability found in its administration to several iterations of the same course. The obvious advantage in the self-assessment approach is a great savings in administration time, not to mention the cost avoidance of having to develop a separate learning evaluation instrument for each distance learning event.

Some of the analyses reported here address the interactions between instruction, management, and technology (or logistics) as developed in Harrison et. al. (1991). Although the management issue was not assessed (because the National Guard Bureau was addressing that separately through "after action reviews"), interactions between instruction and technology were presented in numerous cross tabulations. However, even though quality of the technology was rated lower by those participants who traveled more than one hour to the training, the amount they learned was equal to those traveling less than an hour. Perhaps this is due to the nature of military training which focuses the students on completion of assignment despite perceptions of less than favorable conditions.

The one-page format met with no resistance from students or site administrators, in fact they seemed to appreciate its brevity. Advanced mailing of sufficient copies of the form along with a letter of explanation and with pre-paid and pre-addressed return envelopes assured a high and rapid return rate. The scannable format allowed analysts to provide quick results to National Guard Bureau training officers, who found external verification that students where learning as useful feedback to justify the continued delivery of such courses. Problems identified with the technical qualities at certain sites, such as poor audio or video quality, usually could be traced to

a satellite or local connectivity problem. A decrease in such problems over time are indicative of a "learning curve" for the site facilitators and program providers.

Another area of interest was the travel time variable. Since the National Guard Bureau is interested in a one-hour or less travel time by its soldiers to a distance learning facility, a baseline for the current training year of 81% of the soldiers who participated in distance learning events can serve as a useful yardstick to measure progress in achieving this goal. Caution should be exercised with this figure since it was based only on those soldiers attending the events reported here.

Although the compressed form is only one page in length, the opportunities to cross tabulate on the basis of demographic factors can lead to insights and trends not conceived in the original design. Such analyses can provide useful feedback to the stakeholders in distance learning, including organizers, managers, instructors, and technicians

This instrument is useful for civilian audiences as well. Of course, there would need to be a few adjustments made, such as excluding rank, but the bulk of the items could be retained. Based on the lessons learned in the present evaluation, a revised version recommended for future use is included in Appendix B. An adjusted version for civilian use is offered in Appendix C.

References

Altschuld, J., & Lower, M. (1984). Improving mailed questionnaires: Analysis of a 96 percent return rate. In D.C. Lockhart (Vol. Ed.), New directions in program evaluation: Vol. 21. Making effective use of mailed questionnaires (pp. 5-18). San Francisco: Jossey-Bass.

Anderson, C., Warner, J., Spencer, C. (1984). Inflation bias in self-assessment examinations: Implications for valid employee selection. Journal of Applied Psychology, 69(4), 574-580.

Baumgartner, R., & Heberlein, T., (1984). Recent research on mailed questionnaire response rates. In D.C. Lockhart (Vol. Ed.), New directions in program evaluation: Vol. 21. Making effective use of mailed questionnaires (pp. 65-76). San Francisco: Jossey-Bass.

Biner, P., Dean, R., & Mellinger, A. (1994). Factors underlying distance learner satisfaction with televised college-level courses. The American Journal of Distance Education, 7 (1), 60-71.

Babbie, E. (1990). Survey Research Methods (2nd Ed.). Belmont, CA: Wadsworth Publishing.

Baker, E.L., & O'Neil, H.F. (Eds.). (1994). Technology assessment in education and training. Hillsdale, NJ: Lawrence Erlbaum Associates.

Bramble, N. & Martin, B. (1995). The Florida teletraining project: military training via two-way compressed video. American Journal of Distance Education, 9(1), 6-26.

Brown, S., Nathenson, M., & Kirkup, G. (1982). Learning from evaluation at the Open University: II. Helping student to lean from audio-visual media. British Journal of Educational Technology, 13 (3), 217-236.

Clark, R. (1994). Assessment of distance learning technology. In E.L. Baker & H.F. O'Neil, Jr. (Eds.), Technology assessment in education and training. (pp. 63-78). Hillsdale, NJ: Lawrence Erlbaum Associates.

Curnow, C. (1998). Self-assessment as a training needs assessment tool. (Technical Report in preparation.) Alexandria, VA. U.S. Army Research Institute for the Behavioral and Social Sciences.

DeNisi, A. & Shaw, J. (1977). Investigation of the uses of self-reports of abilities. Journal of Applied Psychology, 62(5), 641-644.

Eichner, K. & Habermehl, W., (1981). Predicting response rates to mailed questionnaires (Comment on Heberlein and Baumgartner, ASR, August, 1978). American Sociological Review, 46, 361-363.

Ellis, J., Knirk, F., Taylor, B., & McDonald, B. (1987). The course evaluation system. (NPRDC-TR-87-19, AD A178 521) San Diego, CA: Navy Personnel Research and Development Center.

Farh, J. & Dobbins, G. (1989). Effects of comparative performance information on the accuracy of self-ratings and agreement between self- and supervisor ratings. Journal of Applied Psychology, 74(4), 606-610.

Ghodsian, D., Bjork, R., & Benjamin, A. (1997). Evaluating training during training: Obstacles and opportunities. In M. Quinones & A. Ehrenstein (Eds.) Training for a rapidly changing workplace. (pp. 63-88). Washington, DC: APA.

Grum, A.F., Campbell, C.C., Montgomery, A.B., Shields, J.L., Thomas, M.U., Vasak, J.T. & Weinberger, P.J. (1995). Army Science Board Ad Hoc Study, "Use of Technologies in Education and Training", Washington, DC: Army Science Board (AD A303 504).

Goyder, J. (1982). Further evidence on factors affecting response rates to mailed questionnaires. American Sociological Review, 47, 550-553.

Harris, M. & Shaubroeck, J. (1988).A meta-analysis of self-supervisor, self-peer, and peer-supervisor ratings. Personnel Psychology, 41, 43-62.

Harrison, P., Seeman, B., & Behm, R. (1991) Development of a distance education assessment instrument. Educational Technology Research and Development, 39, 65-77

Howard, F. (1997). Distance Learning Annotated Bibliography. (TRAC-WSME-TR-97-015, AD A330 045). White Sands Missle Range, NM. TRADOC Analysis Center.

Jegede, O., Gooley, A., Towers, S. (1997). An evaluation of the Queensland open learning network audiographic conferencing professional development programs. Available HTTP: www.usq.edu.au/electpub/e-jist/vol1no4/jegede.htm

Jones, T. (1992) . IITS Students' evaluation questionnaire for the fall semester 1991. A Summary Report. (Eric Document Reproduction Services No. ED 345 716).

Kabat, E. & Freidel, J. (1990). The development , pilot-testing, and dissemination of a comprehensive evaluation model for assessing the effectiveness of a two-way interactive distance learning system. Final report. Eastern Iowa Community college district office of academic affairs and planning.

Kanuk, L. & Berenson, C. (1975) Mail surveys and response rates: A literature review. Journal of Marketing Research, 12, 440-453

Lehman, L.A. & Kinney, P.A. (1992). Distance learning pilot: video teletraining reserve component (VTT-RC) training effectiveness analysis (TRAC-WSMR-TEA-92-0105, AD B196 594). White Sands Missile Range, NM: TRADOC Analysis Center.

Levine, E., Flory, A., Ash, R. (1977). Self-assessment in personnel selection. Journal of Applied Psychology, 62(4), 428-435.

Mackin, D. & Hoffman, J. (1996). Evaluation that can be conducted at a distance: Focus groups over interactive television (ITVO and more. D.O.E. Safeguards & Security Central Training Academy.

Meyer, H. (1980). Self-appraisal of job performance. Personnel Psychology, 33, 291-295.

Moore, M.G. and Kearsely, G. (1996) Distance Education: A Systems View. Belmont, CA: Wadsworth.

National Guard Bureau (1996). Distributive Training Technology Project: Site Planning and Preparation Guide. (NGB-AIS-T, October 10, 1996). Arlington, VA: National Guard Bureau.

OTA (1989). "Linking for learning" A new course for education. U.S. Congress Office of Technology Assessment (OTA-SET-430). Washington, DC: U.S. Government Printing Office.

Phelps, R.R., Wells, R., Ashworth, B., & Hahn, H. (1991). Effectiveness and costs of distance education using computer-mediated communication. American Journal of Distance Education, 5(3), 7-19.

Roszkowski, M.J. & Bean, A.G. (1990) Believe it or not! Longer questionnaires have lower response rates. Journal of Business and Psychology, 4, 495-509.

Russell, T.L. (1996). The "no significant difference" phenomenon. [Online] Available: http://tenb.mta.ca/pheno/phenom.html.

Schendel, J. Morey, J. Grainier, Hall, S. (1983). Use of self assessment in estimating levels of skill retention. (Research Report 1341, AD A141 042). Alexandria, VA: U.S. Army Research Institute for the Behavioral and Social Sciences.

Seidel, R. J., & Chatelier, P.R. (Eds.). (1994). Learning Without Boundaries: Technology to Support Distance/Distributed Learning. New York: Plenum Press.

Simpson, H., Wetzel, D., & Pugh, L. (1995). Delivery of division officer Navy leadership training by videoteletraining: Initial concept and test evaluation (NPRDC-TR-95-7, AD A298 102) San Diego, CA: Navy Personnel Research and Development Center.

U.S. Army Training Doctrine Command. (1996). Army Distance Learning Program (Operations Directive No. I-96). Fort Monroe, VA: Author.

Wetzel, D. (1995). Evaluation of a celestial navigation refresher course delivered by videoteletraining. (AD A300 925) San Diego, CA: Navy Personnel Research and Development Center.

Wetzel, D., Radtke, P., Parchman, S., & Seymour, G. (1996). Delivery of a fiber optic cable repair course by videoteletraining (NPRDC- TR-96-4, AD A304 318). San Diego, CA: Navy Personnel Research and Development Center.

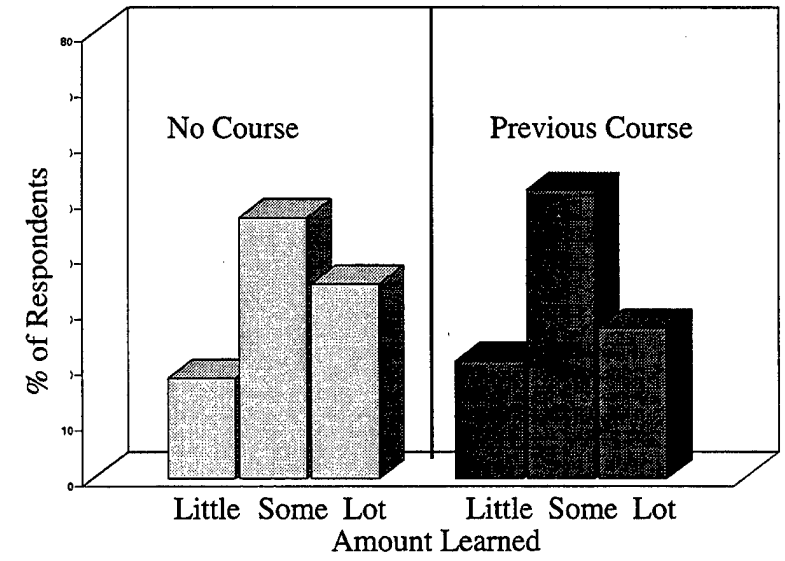
Wetzel, D., Simpson, H., & Seymour, G., (1995). The use of videoteletraining to deliver chief and leading petty officer Navy leadership training: Evaluation and summary (AD A298 374). San Diego, CA: Navy Personnel Research.

Wexley, K.N. & Lathan, G.P. (1981) Developing and Training Human Resources in Organizations. Glenview, IL: Scott, Foresman and Company

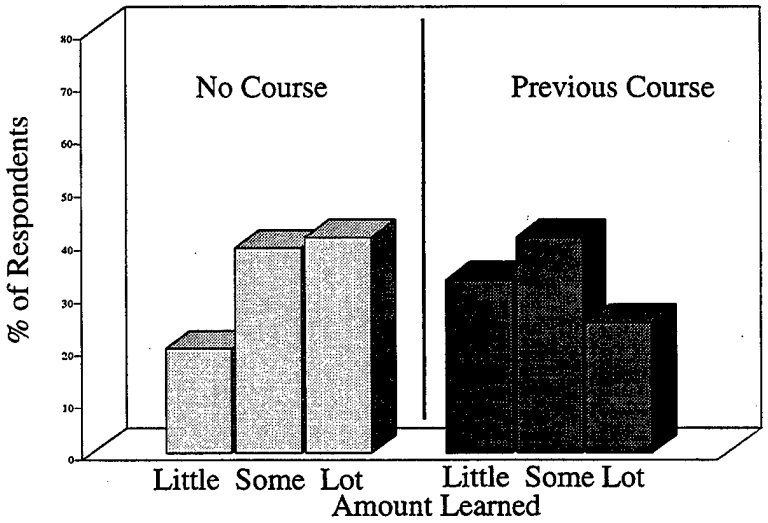
Wisher, R.A., Priest, A.N. & Glover, E. (1997) Audio Teletraining for unit clerks: a cost-effectiveness analysis. (Research Report 1712, AD A337 689) Alexandria, VA: U.S. Army Research Institute for the Behavioral and Social Sciences.

Wisher, R., Seidel, R., Priest, A., Knott, B., Curnow, C. (1997, March) Distance learning over extended periods: The effects of knowledge decay. Paper presented at the annual conference of the American Educational Research Association, Chicago, IL.

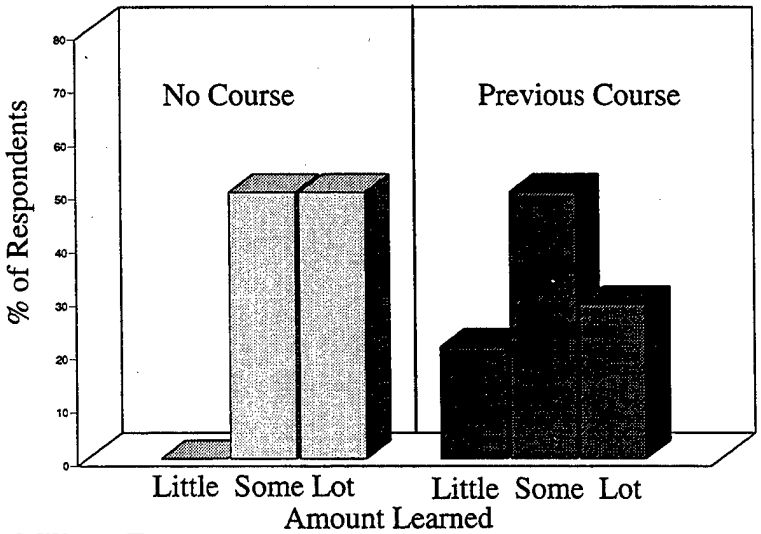
Yammarino, F.J., Skinner, S.J., Childers, T.L. (1991). Understanding mail survey response behavior: A meta-analysis. Public Opinion Quarterly, 55, 613-639.



Airborne Call for Fire (n=87)

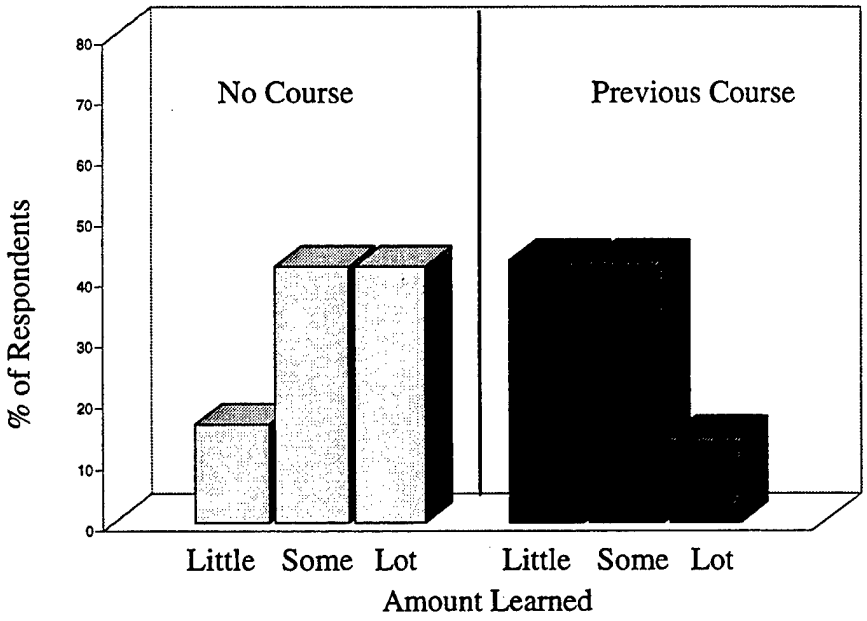


Terrorism Update (n=297)

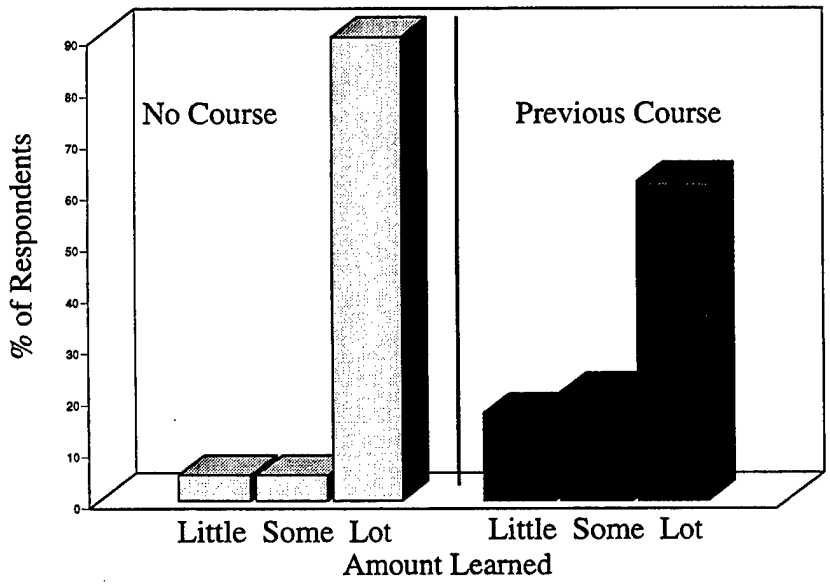


Airspace Management (n=18)

Figure 4. Other Military Events



Creative Problem Solving (n=21)



Counter Drug Event (n=72)

Figure 5. Other Civilian Events

Appendix A
Original Evaluation Form

Distance Learning Evaluation Sheet (Original)

Location of training _____ _____ Today's Date: ____/____/____
City State Day Month Year

1. What is your rank?

- E1 toE4
- E5 to E9
- Warrant Officer
- 2LT, 1LT, or CPT
- MAJ, LTC, COL or higher
- Other _____

2. Mark the highest level of education you have attained:

- High School Diploma
- GED
- Some College, No Degree
- Associate Degree
- Bachelor's Degree
- Advanced Degree

3. How old were you on your last birthday? Yrs.

4. Have you ever completed a full tour (2 or more years) of Active Duty in the Regular Army?
yes no

5. Was this a live event or a videotaped version? Mark One.
Live Videotaped

6. How important did you feel it was for you to attend this training? Respond using the five point scale below.
Not important at all 1 2 3 4 5 Very important

7. Have you had previous training related to today's training topic?
Yes No

8. Compared to what you already knew about topics related to today's training, how much more did you learn in this training event? Respond using the 5 point scale below.
None 1 2 3 4 5 A Lot More

9. About how many other interactive TV training events similar to this have you previously participated in?

	None	1 or 2	3 or 4	5 to 10	10 or more
For Guard/Reserve training?					
For your civilian job?					
For education or occupational training?					
For other interests (e.g. hobbies etc.)?					

10. Please respond to the following questions using a five point scale with 1 being "Poor" and 5 being "Excellent." From your vantage point, how would you rate:

	Poor 1	2	3	4	Excellent 5	Not Applicable
Location of the video screen:						
Quality of audio:						
Quality of video:						
Instructor effectiveness:						
Opportunity to ask questions:						
Responsiveness to student questions:						
Relevance of course to Guard duties/job:						
Overall learning environment (i.e., lighting, distractions, room size, etc.):						
Overall effectiveness of instruction:						

11. How long did it take you to travel to this training facility?
1 to 29 min. 30-59 min. 1-1 1/2 hours 1 1/2 hours or more

This information is being collected by the U. S. Army Research Institute in accordance with the Privacy Act of 1974. Information on individuals is confidential and will not be released to anyone. Group information will be used only for research and policy analysis.

Appendix B
Revised Evaluation Form for the Military

NGB Distance Learning - Evaluation Sheet

Location of training _____ (City) _____ (State) Today's Date: ____/____/____ (Day Month Year)

1. What is your rank?

- ○ E1 toE4
- ○ E5 to E9
- ○ Warrant Officer
- ○ 2LT, 1LT, or CPT
- ○ MAJ, LTC, COL or higher
- ○ Other _____

2. Was this a live event or a videotaped version? Mark One.

○ Live ○ Videotaped

3. How important did you feel it was for you to attend this training? Respond using the five point scale below.

Not important at all ○ ○ ○ ○ ○ Very important

4. Have you had previous training related to today's training topic?

Yes ○ No ○

5. Compared to what you already knew about topics related to today's training, how much more did you learn in this training event? Respond using the 5 point scale below.

None ○ ○ ○ ○ ○ A Lot More

6. About how many other interactive TV training events similar to this have you previously participated in?

	None	1 or 2	3 or 4	5 to 10	10 or more
For Guard/Reserve training?	○	○	○	○	○
For your civilian job?	○	○	○	○	○
For education or occupational training?	○	○	○	○	○
For other interests (e.g. hobbies etc.)?	○	○	○	○	○

7. Please respond to the following questions using a five point scale with 1 being "Poor" and 5 being "Excellent." From your vantage point, how would you rate:

	Poor				Excellent	Not Applicable
Location of the video screen:	○	○	○	○	○	○
Quality of audio:	○	○	○	○	○	○
Quality of video:	○	○	○	○	○	○
Instructor effectiveness:	○	○	○	○	○	○
Opportunity to ask questions:	○	○	○	○	○	○
Responsiveness to student questions:	○	○	○	○	○	○
Relevance of course to Guard duties/job:	○	○	○	○	○	○
Overall learning environment (i.e., lighting, distractions, room size, etc.):	○	○	○	○	○	○
Overall effectiveness of instruction:	○	○	○	○	○	○

8. How long did it take you to travel to this training facility?

○ 1 to 29 min. ○ 30-59 min. ○ 1-1 1/2 hours ○ 1 1/2 hours or more

This information is being collected in accordance with the Privacy Act of 1974. Information on individuals is confidential and will not be released to anyone. Group information will be used only for research and policy analysis.

Appendix C
Revised Evaluation Form for General Use

Distance Learning Evaluation Sheet (Civilian)

Location of training _____ (City) _____ (State) Today's Date: ____/____/____ (Day Month Year)

1. How important did you feel it was for you to attend this training? Respond using the five point scale below.

   Not important at all ○ ○ ○ ○ ○ Very important

2. Have you had previous training related to today's training topic?

   Yes ○ No ○

3. Compared to what you already knew about topics related to today's training, how much more did you learn in this training event? Respond using the 5 point scale below.

   None ○ ○ ○ ○ ○ A Lot More

4. About how many other interactive TV training events similar to this have you previously participated in?

	None	1 or 2	3 or 4	5 to 10	10 or more
For your civilian job?	○	○	○	○	○
For education or occupational training?	○	○	○	○	○
For other interests (e.g. hobbies etc.)?	○	○	○	○	○

5. Please respond to the following questions using a five point scale with 1 being "Poor" and 5 being "Excellent." From your vantage point, how would you rate:

	Poor				Excellent	Not Applicable
Location of the video screen:	○	○	○	○	○	○
Quality of audio:	○	○	○	○	○	○
Quality of video:	○	○	○	○	○	○
Instructor effectiveness:	○	○	○	○	○	○
Opportunity to ask questions:	○	○	○	○	○	○
Responsiveness to student questions:	○	○	○	○	○	○
Relevance of course to job:	○	○	○	○	○	○
Overall learning environment (i.e., lighting, distractions, room size, etc.):	○	○	○	○	○	○
Overall effectiveness of instruction:	○	○	○	○	○	○